УДК 004.056:336.71(076)

ИССЛЕДОВАНИЕ ВЛИЯНИЯ СЕМАНТИЧЕСКИХ ИЗМЕНЕНИЙ ОБЕЗЛИЧЕННЫХ ПЕРСОНАЛЬНЫХ ДАННЫХ НА ИХ ИНФОРМАЦИОННУЮ БЕЗОПАСНОСТЬ

А.М. Тимофеев[1], К.Р. Восковцева[2], Я.А. Клиндухов[2]

[1]Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники», Минск, Беларусь

[2]Учреждение образования «Национальный детский технопарк», Минск, Беларусь

Аннотация. Предложены принципы реализации метода изменения состава или семантики, заключающиеся в использовании наборов различных доверительных вычислительных баз (ДВБ). Такие принципы соответствуют требованиям законодательства Республики Беларусь в сфере защиты информации и могут быть использованы при проектировании, создании и эксплуатации (модернизации) современных информационных систем, посредством которых осуществляется автоматизированная обработка биометрических, генетических либо специальных персональных данных. Выполнены исследования влияния семантических изменений обезличенных персональных данных на их информационную безопасность, и получена зависимость вероятностей появления обезличенных символов персональных данных от их номеров, содержащихся в ДВБ. Установлено, что с ростом количества ДВБ и с увеличением количества исходных и соответствующих им обезличенных персональных данных, содержащихся в каждой такой ДВБ, отклонение вероятности появления исходных

персональных данных от вероятности появления обезличенных персональных данных проявляется в бóльшей мере.
Ключевые слова: информационные системы; персональные данные; защита информации; деобезличивание персональных данных; методы обезличивания персональных данных; метод изменения состава или семантики.

STUDY OF THE IMPACT OF SEMANTIC CHANGES OF DEPERSONALIZED PERSONAL DATA ON THEIR INFORMATION SECURITY

[1]A. Timofeev, [2]K. Voskovtseva, [2]Y. Klindukhov
[1]Education Institution "Belarusian State University of Informatics and Radioelectronics", Minsk, Belarus
[2]Education Institution "National Children's Technopark", Minsk, Belarus

Abstract. The principles of implementing the method of changing the composition or semantics are proposed. These principles consist in using sets of different trusted computing bases (TCB). This approach complies with the legislation of the Republic of Belarus in the field of information security. The implementation of the method of changing the composition or semantics can be used in the design, creation and operation (modernization) of modern information systems that are used for automated processing of biometric, genetic or special personal data. The influence of semantic changes in depersonalized personal data on their information security has been studied. The dependence of the probabilities of occurrence of depersonalized symbols of personal data on their numbers contained in the TCB has been obtained. It has been established that with an increase in the number of TCBs and with an increase in the number of original and corresponding depersonalized personal data contained in each such TCB, the deviation of the probability of occurrence of the original personal data from the probability of occurrence of depersonalized personal data manifests itself to a greater extent.
Keywords: information systems; personal data; information security; depersonalization of personal data; methods of depersonalization of personal data; method of changing the composition or semantics.

Введение

В настоящее время одной из наиболее важной задач, решаемых при построении информационных систем типовых классов, является обеспечение их информационной безопасности [1–3]. При этом важно учитывать наличие в таких системах любых персональных данных, кроме общедоступных.

Персональными данными называют любую информацию, относящуюся к идентифицированному физическому лицу или физическому лицу, которое может быть идентифицировано.

Под общедоступными персональными данными будем понимать персональные данные, распространенные либо самим субъектом персональных данных, либо с его согласия или распространенные в соответствии с требованиями законодательных актов Республики Беларусь.

В соответствии с требованиями законодательства Республики Беларусь в сфере защиты информации для обеспечения информационной безопасности персональных данных выполняют их обезличивание с использованием методов, определенных в Приказе Оперативно-аналитического центра при Президенте Республики Беларусь от 10 декабря 2024 г. №259 «Об изменении приказов Оперативно-аналитического центра при Президенте Республики Беларусь от 28 марта 2014 г. № 26 и от 20 февраля 2020 г. № 66». Процедура обезличивания персональных данных подразумевает действия, в результате которых становится невозможным без использования дополнительной информации определить принадлежность персональных данных конкретному субъекту персональных данных.

Субъект персональных данных – это физическое лицо, в отношении которого осуществляется обработка персональных данных

Одним из таких методов является метод изменения состава или семантики, сущность которого заключается в том, что выполняют обобщение, изменение или удаление части сведений, позволяющих идентифицировать субъекта персональных данных. При этом полученные обезличенные персональные данные и правила их изменения необходимо хранить раздельно.

Отметим, что в случае обобщения или удаления части исходных персональных данных при реализации процедуры их обезличивания не выполняется свойство полноты обезличенных персональных данных. Это не позволит выполнить деобезличивание персональных данных без использования соответствующих таблиц. Применение указанных таблиц создает уязвимость информационных систем, что является недостатком метода изменения состава или семантики. От этого недостатка свободные методы [4], которые предусматривают дополнительно изменение семантики обезличенных персональных данных по отношению к семантике исходных персональных данных. Однако указанная замена сохраняет статистику естественного языка, что может быть использовано нарушителем для доступа к исходным персональным данным и реализовано на основе методов частотного анализа [5]. В связи с этим целью данной работы являлась реализация семантических изменений исходных персональных данных для решения задач их обезличивания, при которых статистические распределения вероятностей появления отдельных обезличенных символов не соответствуют статистическим распределениям вероятностей появления символов исходных персональных данных.

Объектом исследования являлся метод изменения состава, применяемый для обезличивания персональных данных. Этот метод выбран в качестве объекта исследования, поскольку он является одним из обязательных методов обезличивания персональных данных для организаций и предприятий, в которых осуществляется обработка биометрических, генетических либо специальных персональных данных с использованием типовых информационных систем в соответствии с требованиями законодательства Республики Беларусь в сфере защиты информации.

Предметом исследования являлось применение набора доверительных вычислительных баз для реализации обезличивания персональных данных на основе метода изменения состава или семантики, при котором статистические распределения вероятностей появления отдельных обезличенных символов не соответствуют статистическим распределениям вероятностей появления символов исходных персональных данных.

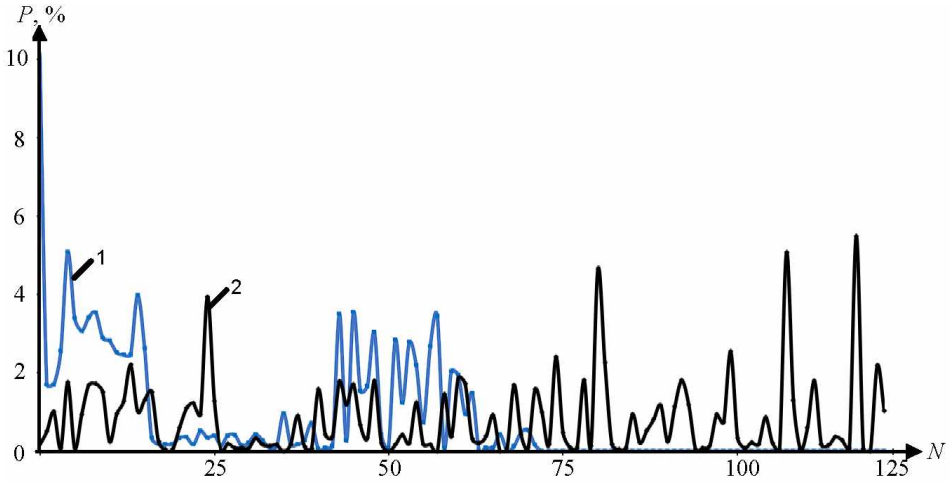
Обезличивание персональных данных с использованием блочной замены и набора различных ДВБ

В рамках данной работы предложена реализация метода изменения состава или семантики, которая заключается в следующем. Персональные данные, подлежащие обезличиванию, вначале разбивают на блоки длинной n символов каждый. В случае если общее число символов не кратно n, то последний блок дополняют символами пробела. Затем формируют n-ое количество доверительных вычислительных баз, которые представляют собой перестановочные таблицы, являются ключевыми элементами информационных систем и хранятся защищенным образом отдельно от исходных и обезличенных персональных данных. После этого первый символ первого блока исходных персональных данных заменяют на обезличенный символ из первой ДВБ, второй символ – из второй ДВБ и т.д. Последний символ первого блока исходных персональных данных заменяют на обезличенный символ из n-ой ДВБ. Второй

и последующий блоки исходных персональных данных обезличивают аналогичным образом с использованием ДВБ $1 \div n$ соответственно.

Отметим, что целесообразно устанавливать общее количество записей, содержащихся в каждой ДВБ, исходя из требований информационной безопасности, которые определяют собственники информационных систем либо до начала проектирования информационных систем, либо на этапе их эксплуатации (модернизации).

Выполнены исследования влияния семантических изменений обезличенных персональных данных на их информационную безопасность, и получена зависимость вероятностей появления обезличенных символов персональных данных от их номеров, содержащихся в ДВБ, которая представлена на рисунке.



Зависимость вероятности появления обезличенных символов персональных данных от их номера из ДВБ. Реализация метода состава или семантики с использованием: 1 – простой замены на базе одной ДВБ; 2 – сложной замены на базе двух ДВБ

Важно отметить, что представленные на рис. результаты получены с использованием двух ДВБ, каждая из которых имеет 780 записей исходных и соответствующих им символов обезличенных персональных данных.

Из приведенных результатов видно, что вероятности появления символов исходных персональных данных в случае использования сложной замены на базе двух ДВБ не соответствуют вероятностям появления обезличенных символов, что наблюдалось при использовании простой замены на базе одной ДВБ. Это позволяет повысить уровень информационной безопасности персональных данных с использованием их обезличивания на основе метода изменения состава или семантики при выполнении сложной замены на базе набора ДВБ за счет того, что частотный анализ обезличенных символов усложняется, в сравнении с частотным анализом символов, обезличенных на базе одной ДВБ.

Выполненная оценка показала, что с ростом количества ДВБ и с увеличением количества исходных и соответствующих им обезличенных персональных данных, содержащихся в каждой такой ДВБ, отклонение вероятности появления исходных персональных данных от вероятности появления обезличенных персональных данных проявляется в бóльшей мере.

Предложенные в данной работе принципы обезличивания персональных данных соответствуют требованиям законодательства Республики Беларусь в сфере защиты информации, реализованы на базе программной платформы Windows Forms на языке программирования высокого уровня C#12 (.NET Framework 4.8) и выполнены в виде

исследовательского проекта учреждения образования «Национальный детский технопарк». Для проведения исследований использовалась специально сгенерированная база исходных персональных данных, полученная с помощью генератора псевдослучайных данных, который выполнен программно (любые совпадения с персональными данными реальных физических лиц случайны; авторы работы и издательство не несут ответственности и не предоставляют гарантий в связи с публикацией в настоящей статье любой информации, относящейся к реальному физическому лицу). Так, например, запись исходных персональных данных №5034 «Сидоров Александр Андреевич» после реализации обезличивания с использованием предложенных в настоящей работе принципов имеет вид «ʢœãн΄ȣÚPɼǴøÐøUÚã΄Ȟ<Ủʱ΄aœŪaʢ».

Важно отметить, что представленные в данной работе принципы обезличивания персональных данных не требуют больших вычислительных ресурсов от современных аппаратно-программных комплексов и характеризуются высоким уровнем информационной безопасности, достаточным для решения практических задач по обезличиванию персональных данных.

Заключение

Предложены принципы реализации метода изменения состава или семантики, позволяющие повысить уровень информационной безопасности информационных систем, в которых обрабатываются персональные данные, за счет использования наборов различных доверительных вычислительных баз. Эти принципы соответствуют требованиям законодательства Республики Беларусь в сфере защиты информации и могут быть использованы при проектировании, создании и эксплуатации (модернизации) современных информационных систем, в частности, для информационных систем типовых классов.

Применение наборов различных доверительных вычислительных баз для решения задач обезличивания персональных данных на основе метода изменения состава или семантики позволило сохранить вычислительную сложность указанного метода и при этом повысить уровень информационной безопасности обезличенных персональных данных за счет усложнения процедуры частотного анализа, который может быть применен возможным нарушителем информационной безопасности.

Список использованных источников

1. Ворона, В. А. (2023) *Биометрическая идентификация личности*. Москва, Горячая линия-Телеком.

2. Коллинз, М. (2020) *Защита сетей. Подход на основе анализа данных*. Москва, ДМК Пресс.

3. Остапенко, Г. А. (2020) *Информационные операции и атаки в социотехнических системах: организационно-правовые аспекты противодействия*. Москва, Горячая линия-Телеком.

4. Солдатова В. И. (2020) Защита персональных данных в условиях применения цифровых технологий. *Lex russica*. (2), 33–43.

5. Арьков, В. Ю. (2023) *Частотный анализ числовых и текстовых данных*. Екатеринбург, Издательские решения.

References

1. Vorona V. A. (2023) *Biometric Identification of Personality*. Moscow, Goryachaya Liniya-Telecom (in Russian).

2. Collins M. (2020) *A Data-Based Approach*. Moscow, DMK Press (in Russian).

3. Ostapenko G. A. (2020) *Information Operations and Attacks in Socio-Technical Systems: Organizational and Legal Aspects of Counteraction*. Moscow, DMK Press (in Russian).

4. Soldatova V. I. (2020) Protection of Personal Data in the Context of Digital Technologies. *Lex russica*. (2). 33–43 (in Russian).

5. Arkov V. Yu. (2023) *Frequency Analysis of Numerical and Text Data*. Ekaterinburg. Publishing Solutions (in Russian).

Сведения об авторах

Тимофеев А.М., канд. техн. наук, доц., доц. каф. защиты информации, учреждение образования «Белорусский государственный университет информатики и радиоэлектроники», tamvks@mail.ru.
Восковцева К.Р., учащаяся, учреждение образования «Национальный детский технопарк», kristiza2009@gmail.com.
Клиндухов Я. А., учащийся, учреждение образования «Национальный детский технопарк», klinyarik1@gmail.com.

Information about the authors

Timofeev A., Cand. Sci. (Tech.). Assosiate Professor, Associate Professor of the Department of Information Protection, Educational Institution "Belarusian State University of Informatics and Radioelectronics", tamvks@mail.ru.
Voskovtseva K., Student, Educational Institution "National Children's Technopark", kristiza2009@gmail.com.
Klindukhov Y., Student, Educational Institution "National Children's Technopark", klinyarik1@gmail.com.